

EvalGRAPHS: herramienta para implementar evaluadores inteligentes. **Metodología de estudio del comportamiento de los evaluadores**

**M. Gloria Sanchez-Torrubia¹, Carmen Torres-Blanc¹,
Silvia García Guerrero¹, Ruben Gonzalez-Martin¹**

¹ *Universidad Politécnica de Madrid*

gsanchez@fi.upm.es, ctorres@fi.upm.es,
silvis.garcia@gmail.com, rubengonzmartin@gmail.com

El Modelo Granular Lingüístico de la Evaluación del aprendizaje humano [1] ha sido desarrollado a partir del paradigma *GLMP* [1], basado en la computación con palabras y percepciones, con el objetivo de diseñar un modelo teórico de representación del aprendizaje que pueda aplicarse a la creación de sistemas (*GLMPs*) que emulen la evaluación realizada por un profesor, en base a sus criterios, y expresen dicha evaluación mediante un informe en lenguaje natural. Un *GLMP* es una estructura jerárquica que organiza y procesa datos mediante inferencia borrosa y modela un fenómeno mediante percepciones o unidades de conocimiento. Estos sistemas utilizan como datos de entrada los generados automáticamente por un entorno de aprendizaje y producen como salida un informe de evaluación que incluye una calificación numérica. De esta manera se puede compaginar la evaluación formativa, comparando logros y objetivos de aprendizaje, que realiza un profesor, con la rapidez y eficacia de un sistema automatizado. Como aplicación del modelo general, hemos diseñado varios sistemas que evalúan, en base a criterios de un profesor, unos los *quizzes* de *Moodle* y otros las simulaciones visuales de algoritmos de grafos hechas en el entorno de aprendizaje *GRAPHS* [2].

Relacionado con el entorno *GRAPHS* hemos desarrollado una herramienta, a la que hemos llamado *EvalGRAPHS*, que ayuda en el diseño y lleva a cabo la ejecución de estos últimos sistemas. La aplicación está compuesta por dos partes: la primera dedicada a la configuración, la salvaguarda, la recuperación y la modificación de los evaluadores que serán implementados por el sistema, y la segunda dedicada a la evaluación, mediante uno de los evaluadores previamente configurados, a partir de los *log* de interacción generados por *GRAPHS* al realizar la simulación de un algoritmo [2]. Dicha evaluación se expresa mediante un informe de evaluación formativa, en lenguaje natural, que describe el nivel de logro del alumno en cada uno de los objetivos de aprendizaje fijados por el profesor y, además, la calificación numérica de dicha simulación.

Cuando se crea una nueva configuración, la herramienta genera un archivo *XML* que almacena todos los datos que la describen: las etiquetas que corresponden a cada uno de los datos de entrada, la estructura de percepciones del *GLMP*, sus etiquetas lingüísticas, las plantillas de generación de texto, las reglas que forman cada uno de los motores de inferencia que obtienen la información de una percepción en función de la de sus subordinadas, la *t-norma* y la *t-conorma* que se utilizan en las funciones de agregación, etc.

Cuando, una vez cargada una configuración concreta para utilizar la herramienta como evaluador, la aplicación pide el *log* de interacción que se desea evaluar, extrae de él los datos de entrada al sistema y genera el informe correspondiente a la simulación evaluada. Además, la aplicación también permite la introducción de datos de forma manual, con el fin de probar las configuraciones que están siendo implementadas.

En este trabajo presentamos una metodología para comprobar la coherencia, el cumplimiento de los objetivos de diseño y el funcionamiento local de un *GLMP*: el estudio de las superficies de comportamiento [3]. Estas superficies representan la salida desborrosificada de una percepción en función de las entradas, también desborrosificadas, a dos de sus percepciones subordinadas. A continuación se describen de forma esquemática los detalles de diseño que se pueden detectar mediante el estudio de estas superficies.

- Influencia de la forma de las etiquetas lingüísticas sobre el funcionamiento del sistema.
 - El conjunto de etiquetas lingüísticas de cada percepción debe ser ortogonal. En caso contrario aparecerán picos y zonas no definidas en las superficies.
 - El uso de etiquetas con forma más *crisp* produce escalones.
- Influencia de la asignación de consecuentes.
 - Como nuestros sistemas se diseñan para la evaluación del aprendizaje, las superficies de comportamiento siempre deben ser crecientes o decrecientes.
 - Crecientes, cuando se agregan aciertos para medir el nivel de acierto o se agregan errores para medir el nivel de error.
 - Decrecientes, cuando se agregan errores para medir nivel de acierto.
 - Las reglas deben ser coherentes, en caso contrario aparecerán zonas de no convexidad.
 - Aunque en nuestros sistemas no pueden existir cambios de comportamiento, en otros *GLMPs* pueden observarse cambios de comportamiento debidos al modelo.
- Validación de que se han definido todas las reglas y usado todas las etiquetas.
 - Si faltan reglas por definir, en la superficie de comportamiento aparecerá una zona no definida (agujero).
 - Si alguna de las etiquetas de la percepción de salida no ha sido asignada en ninguno de los consecuentes de las reglas, aparecerá un escalón en la superficie.

Los sistemas descritos han sido desarrollados, probados según la metodología descrita y mediante la comparación, sobre casos reales, de los resultados generados por el sistema con los facilitados por el profesor cuyos criterios emulan. Además han sido utilizados en el aula con resultados muy satisfactorios.

Referencias

- [1] Sanchez-Torrubia, M. Gloria, Torres-Blanc, Carmen, y Trivino, Gracian: *An approach to automatic learning assessment based on the computational theory of perceptions*. Expert Syst. Appl. **39**, 15 (2012) 12177–12191.
- [2] Sanchez-Torrubia, M. Gloria, Torres-Blanc, Carmen, y Escribano-Blanco, Sonia: *GRAPHS: a learning environment for graph algorithm simulation primed for automatic fuzzy assessment*. En Proc. 10th Koli Calling. (2010) 62–67.
- [3] Sanchez-Torrubia, M. Gloria: *Especificaciones eMathTeacher, creación del Modelo Granular Lingüístico de la evaluación del aprendizaje humano, y diseño de sistemas que implementan estos conceptos*. UPM, Madrid. (2016).